BioHealthBase
BioDefense Public Health Database

The BioHealthBase and Host-Pathogen Interactions

BRC Annual Meeting 3

NIH Contract Number: N01-AI-40041

February 6, 2006

UT SOUTHWESTERN MEDICAL CENTER

NATIONAL INSTITUTES OF HEALTH

NIAID

NORTHROP GRUMMAN
Information Technology

Outline

BioHealthBase BRC Overview
- Pathogens
- System v1.0 Components
- System design

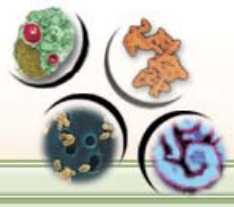
BioHealthBase v1.0 Content - *Francisella tularensis*
- Genome and protein feature annotation
- Pathway hole filling

BioHealthBase v1.0 Content - *Influenza virus*
- Support for comparison of large numbers of related sequences
- Epitope predictions

Host-Pathogen Interactions
- Defining the scope of H-P interactions
- The BioHealthBase/Reactome collaboration for flu
- Future plans

Summary and Future Directions for the BioHealthBase

Acknowledgments



BioHealthBase
BioDefense Public Health Database

BioHealthBase BRC Overview

BioHealthBase Pathogens

Francisella tularensis
Influenza virus
Mycobacterium tuberculosis
Microsporidia
Giardia lamblia
Ricinus communis

- Each pathogen is not only a potential bioterrorism agent, but is also endemic or re-emerging in the United States
  - dual purpose: biodefense and public health
- At least one pathogen from NIAID Category A, B, and C
- Pathogens represent each of the three major classes of microorganism - bacteria, viruses, and parasites - and even plants
- Developing a database that supports these disparate organism types represents a significant challenge. However, the extensible database and system structure would support the addition of other organisms in the future

Initial Pathogen Focus

BioHealthBase
BioDefense Public Health Database

- *Francisella tularensis*
  - Availability of the first genome sequence
  - Early interaction with BioCyc
  - Local expertise
  - Need for Ftu-focused system
- *Influenza virus (types A, B, C)*
  - Avian flu public health concerns
  - Availability of data
  - Different system requirements
- *Mycobacterium tuberculosis*
  - Initial support of a related organism
  - Public health focus

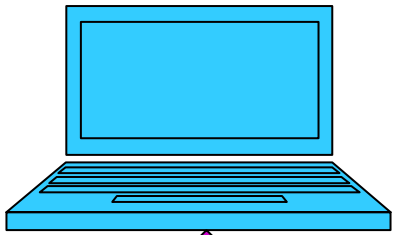
Design Constraints

BioHealthBase
BioDefense Public Health Database

- Support for diverse data types
- Support for diverse data structure
- Support for diverse data availability

System Diagram

Analysis Tools (Future)
- *Blast*
- *Clustal-W/MUSCLE*



Query/Visualization Tools
- GBrowse genomic browser
- Influenza database search
- Bacteria database search
- Gene/protein data display

Flexible, layered Java interface

Data Warehouse (Publication)

Optimized for query access

GUS (Staging)

Data Loading (NCBI, UniProt, PFam, etc.)

Value-added products (polymorphisms, epitopes, protein localization, operons, etc.)

Data validation, data cleansing

v1.0 System Components

BioHealthBase
BioDefense Public Health Database

- BioHealthBase v1.0 deployed on January 31, 2006!
  www.biohealthbase.org
- Database
  - Genome sequence data
  - Gene annotation data
  - Protein annotation data
  - Pathway data
  - Integrated from various sources - NCBI, UniProt, PFam, BioCyc
- Precompiled integrated analysis results (value-added)
  - Influenza consensus sequences and polymorphism (MUSCLE)
  - Influenza MHC class I epitope predictions (NetCTL)
  - Protein localization (PSortB, SignalP, DAS, SubLoc)
  - Operon identification (BioCyc Pathway Tools)
- Query interface
- Genome browser

BHB Home Page

BioHealthBase
BioDefense Public Health Database

HOME | FEEDBACK | ORGANISMS

- Release Notes
- FAQ
- Feedback
- Related Links
- Data Loads

Organisms

- *Influenza Virus*
- *Francisella tularensis*
- *Mycobacterium*

Our Mission

The primary mission of the BioHealthBase system is to assist scientific researchers in their development of vaccines, therapeutics, and diagnostics. The National Institute of Allergy and Infectious Disease (NIAID) Division of Microbiology and Infectious Diseases (DMID) recognizes the challenge posed by bioterrorism, the emergence of disease due to drug-resistant variants of etiologic organisms. DMID has envisioned a consortium of Bioinformatics Resource Centers (BRCs) for Biodefense and Emerging/Re-emerging Infectious Diseases that will provide information technology (IT) support for experimental studies of pathogenic organisms that could be used for biowarfare and bioterrorist activities, many of which also pose an ongoing threat to public health.

< More >

News/Publications/Releases

- Nov 3, 2005 : Began collaboration with Reactome team on interferon pathway analysis
- Nov 6-8, 2005 : Tularemia Workshop, Jiminy Peak Mountain Resort
- Jan 31, 2006 : First production BioHealthBase release
- Feb 6, 2006 : Reactome release with Influenza lifecycle and interferon interference additions
- Feb 6, 2006 - Feb 7, 2006 : 3rd Bioinformatics Resource Centers and Inter-Operability Working Group Meeting, University of Pennsylvania
- Mar 28, 2006 - Apr 2, 2006 : Keystone conference

- Advances in Influenza Research: From Birds to Bench to Bedside (X8)
- Sheraton Steamboat Resort
- Steamboat Springs, Colorado

Genomes in BioHealthBase

Species-based

Species	Kingdom	# Strain
Francisella tularensis	Bacteria	1
Influenza A	Virus	5102
Influenza B	Virus	1040
Influenza C	Virus	152
Mycobacterium avium	Bacteria	1
Mycobacterium bovis	Bacteria	1
Mycobacterium leprae	Bacteria	1
Mycobacterium tuberculosis	Bacteria	2

Kingdom-based

Kingdom	# Species	# Strain
Bacteria	5	6
Virus	3	6294
Total	8	6300

Francisella Home Page

Francisella tularensis

BioHealthBase

HOME | FEEDBACK | OUR MISSION | HELP

Database Search
- Gene
- Locus
- Public Database Identifier
- Gene Ontology

Release Notes

FAQ

Feedback

Related Links

Data Loads

What's New

- Nov 6-8, 2005 : Tularemia Workshop, Jiminy Peak Mountain Resort
- Jan 31, 2006 : First production BioHealthBase release

About *Francisella tularensis*

Francisella tularensis, the causative agent of tularemia, is considered a potential bioterrorist agent. It was named after Dr. Edward Francis and the location where the organism was discovered, Tulare County, California. Important premises are an extremely low infectious dose and a potential for airborne transmission. The two most clinically important entities of tularemia, type A and type B, correspond to the highly virulent subspecies *F. tularensis* subsp. *tularensis* and the moderately virulent *F. tularensis* subsp. *holarctica*, respectively. Taxonomic work has identified two additional subspecies, *F. tularensis* subsp. *mediasiatica*, exhibiting a moderate virulence, and "*F. tularensis* subsp. *novicida*," with a low virulence in animals and humans. The latter subspecies has less-fastidious extracellular growth requirements than the other subspecies and a distinct lipopolysaccharide O-antigen.

Based on small subunity RNA sequences, *F. tularensis* is classified as a member of the gamma-subgroup of proteobacteria. The two species *F. tularensis* and *F. philomiragia* and in addition a number of more recently identified tick endosymbionts are the only members of the genus *Francisella*, which diverges deeply among the gamma-proteobacteria.

< More >

Genome Statistics

General Information for *Francisella tularensis*

Number of Strains in BioHealthBase	1
Genome Type	one circular chromosome
Genome Size	~1800 - 2000 kb
Taxon ID	263
Kingdom	Bacteria

Genome Statistics for Reference Strain *SCHU S4*

		SCHU S4
Length		1892819 bp
GC Content		32%
Taxon ID		177416
Total Genes		1852 (100%)
Protein Coding Genes		1603 (86.6%)
	Genes w/ Database Match	1249 (67.4%)
	Hypothetical Protein	354 (19.1%)
Pseudogenes		201 (10.9%)
tRNA Genes		38 (2.1%)
rRNA Genes		10 (0.5%)

Query Page

BioHealthBase
BioDefense Public Health Database

Gene Search

Text search of genes in BioHealthBase. Leave the term field blank to search for all genes. ?

Specify Organism: Francisella tularensis Schu 4

* Required

Search Type: Gene Symbol

Search Term: panC

Genomic location (optional)

Specify Strand: Both

Specify Start:

Specify End:

GO CLEAR

Gene Summary Page

BioHealthBase
BioDefense Public Health Database

Francisella tularensis Gene Search Results

Your search returned " **1** " records

New Search

GFF submit

	Gene Symbol	Gene Product Name	Pathway Name	Entrez Gene ID	Locus Name	Genbank Genome Accession	UniProtKB Accession	Organism	GBrowse
☐	panC	Pantoate-beta-alanine ligase	aspartate superpathway. et al	3191378	FTT1390	NC_006570	Q5NF57	*Francisella tularensis Schu 4*	View GBrowse

Top

Gene Details Page

BioHealthBase
BioDefense Public Health Database

Gene Details

Strain Identification *1

Organism Name:	Francisella tularensis
Strain Name:	Schu 4
NCBI Taxon ID:	177416

Gene Identification *1

Gene Symbol:	panC
Gene Product Name:	Pantoate-beta-alanine ligase
Locus Name:	FTT1390
Entrez Gene ID:	3191378
Comment:	Similar to AAO89975 (Q83EA3) Pantoate-beta-alanine ligase from Coxiella burnetii (257 aa). FASTA: opt: 692 Z-score: 830.8 E(): 2e-38 Smith-Waterman score: 692; 40.927 identity in 259 aa overlap

Genomic Location *1

Genbank Genome Accession:	NC_006570
Coordinates(5'..3'):	1435067 .. 1435852
Strand:	Forward
Gene Length:	786
Sequence:	View Gene Sequence

Gene Structure

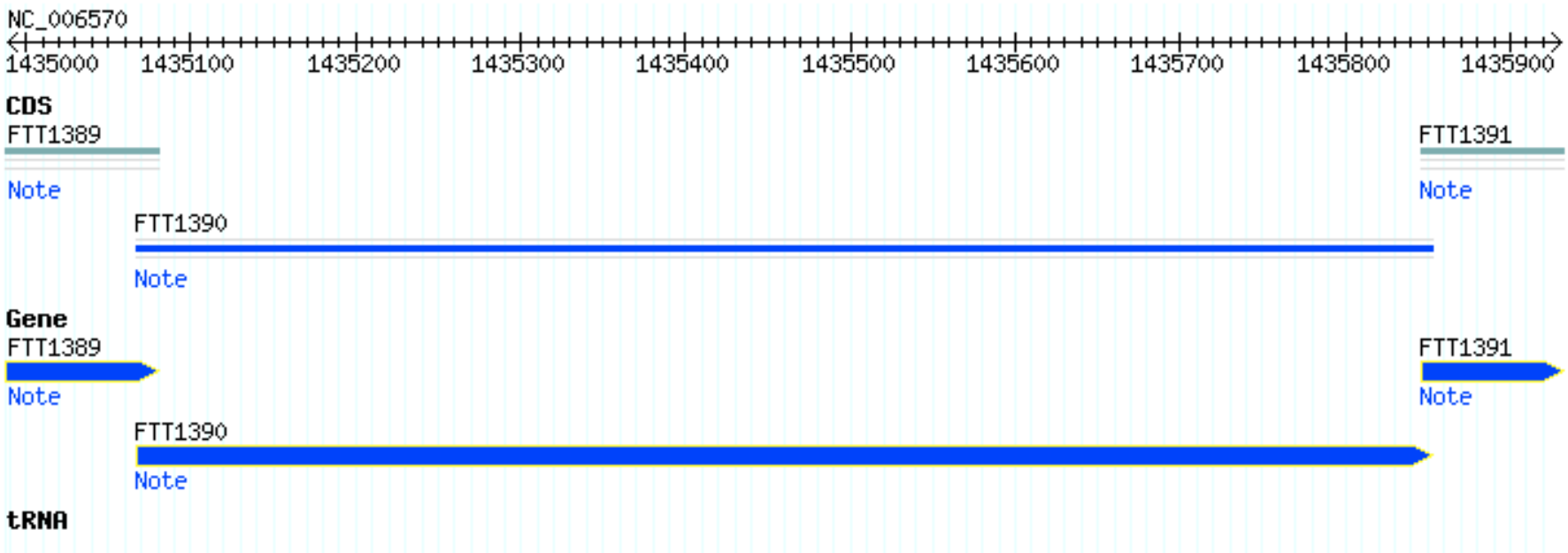
BioHealthBase
BioDefense Public Health Database

Gene Features
*1, *2

CDS	Start	End	Protein Length(aa)	Sequence
1	1435067	1435852	261	Protein

Genomic View
*3

Top



Protein Information

BioHealthBase
BioDefense Public Health Database

Protein Identification *2

Top

Protein Name:	Pantoate-beta-alanine ligase
UniProtKB Accession:	Q5NF57
Genbank Protein Accession:	YP_170335.1
Genbank Protein GI:	56708439
Comment:	Similar to AAO89975 (Q83EA3) Pantoate-beta-alanine ligase from Coxiella burnetii (257 aa). FASTA: opt: 692 Z-score: 830.8 E(): 2e-38 Smith-Waterman score: 692; 40.927 identity in 259 aa overlap
Keywords:	Complete proteome;Ligase

HMM/Pfam Domains *4

Top

Accession	Name	Description	Start	End
PF02569.5	Pantoate_ligase	Pantoate-beta-alanine ligase	1	255

Protein Localization

Top

Location	Raw Score	Range	Program Name
Cytoplasmic	0.097	0-1	SignalP
Cytoplasmic	8.96	0-10	PSORTb
Cytoplasmic	N/A	None	DAS
Cytoplasmic	3.0	0-10	Subloc

Gene Ontology *2

Top

Type	Name	Go ID
FUNCTION	carbonate dehydratase activity	GO:0004089
FUNCTION	ligase activity	GO:0016874
FUNCTION	pantoate-beta-alanine ligase activity	GO:0004592
FUNCTION	zinc ion binding	GO:0008270
PROCESS	one-carbon compound metabolism	GO:0006730
PROCESS	pantothenate biosynthesis	GO:0015940

Pathway/Operon

BioHealthBase
BioDefense Public Health Database

Pathway *5

Top

Pathway Name
pantothenate biosynthesis I
pantothenate and coenzyme A biosynthesis
aspartate superpathway

Operon Identification

Top

Locus Name	Gene Product Name	Score	Range	Method
FTT1388	hypothetical protein	.98	0-1	PTools
FTT1389	3-methyl-2-oxobutanoate hydroxymethyltransferase	.99	0-1	PTools
FTT1390	Pantoate-beta-alanine ligase	.99	0-1	PTools
FTT1391	Aspartate-1-decarboxylase	.81	0-1	PTools
FTT1392	transcriptional regulator	N/A	N/A	PTools

References *1, *2

Top

PubMed ID	Journal Name	Title	Author
15640799	Nat. Genet.	The complete genome sequence of Francisella tularensis, the causative agent of tularemia.	Larsson P. *et al.*

* Data Sources

Top

ID	Source
1	NCBI
2	UniProtKB
3	Gbrowse Tool
4	Pfam
5	BioCyc

Pathway Hole Filling

BioHealthBase
BioDefense Public Health Database

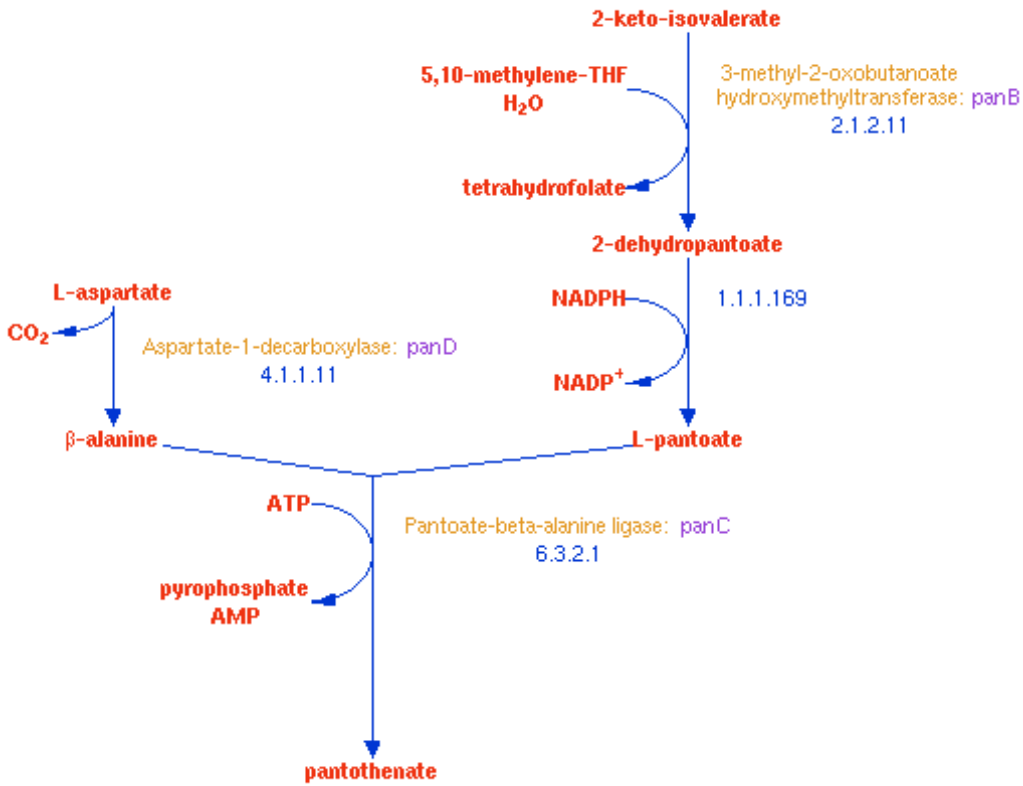
- Semi-automated genome annotation based on sequence alignment with known genes at relatively stringent similarity score cutoff
- For *Francisella tularensis*, ~22% (~350) predicted open reading frames are listed as hypothetical proteins
- Use other orthogonal data to infer relationships to justify use of lower similarity scores
- Metabolic pathway holes and operon prediction relationships

Pantothenate Biosynthesis Pathway

BioHealthBase
BioDefense Public Health Database

F. tularensis Pathway: pantothenate biosynthesis I

More Detail | Less Detail | Cross-Species Comparison



Locations of Mapped Genes:

Pathway Evidence Glyph:

Key to pathway glyph edge colors:

- green: an enzyme catalyzing this reaction is present in this organism
- black: no enzyme catalyzing this reaction has been identified in this organism
- orange: the reaction and any enzyme that catalyzes it (if one has been identified) is unique to this pathway

NADP Oxidoreductase

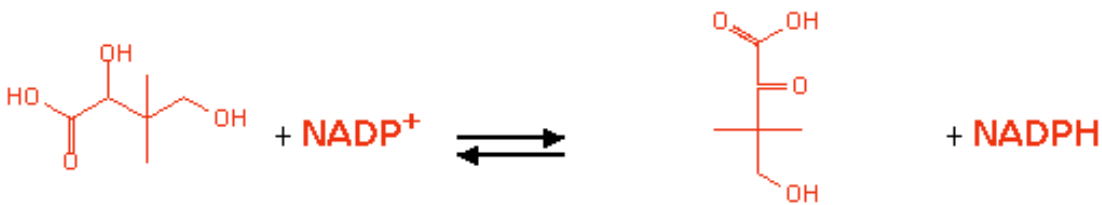
BioHealthBase
BioDefense Public Health Database

F. tularensis Reaction: 1.1.1.169

Cross-Species Comparison

Superclasses: EC-Reactions -> 1 -- Oxidoreductases -> 1.1 -- Acting on the CH-OH group of donors -> 1.1.1 -- With NAD(+) or NADP(+) as acceptor

In Pathway: pantothenate biosynthesis I



L-pantoate

2-dehydropantoate

The reaction direction shown, that is, A + B <==> C + D versus C + D <==> A + B, is in accordance with the Enzyme Commission system.

Unification Links: ENZYME:1.1.1.169

Pathway/Operon

BioHealthBase
BioDefense Public Health Database

Pathway *5

Top

Pathway Name
pantothenate biosynthesis I
pantothenate and coenzyme A biosynthesis
aspartate superpathway

Operon Identification

Top

Locus Name	Gene Product Name	Score	Range	Method
FTT1388	hypothetical protein	.98	0-1	PTools
FTT1389	3-methyl-2-oxobutanoate hydroxymethyltransferase	.99	0-1	PTools
FTT1390	Pantoate-beta-alanine ligase	.99	0-1	PTools
FTT1391	Aspartate-1-decarboxylase	.81	0-1	PTools
FTT1392	transcriptional regulator	N/A	N/A	PTools

References *1, *2

Top

PubMed ID	Journal Name	Title	Author
15640799	Nat. Genet.	The complete genome sequence of Francisella tularensis, the causative agent of tularemia.	Larsson P. *et al.*

* Data Sources

Top

ID	Source
1	NCBI
2	UniProtKB
3	Gbrowse Tool
4	Pfam
5	BioCyc

Genome Browser - panC Operon

BioHealthBase
BioDefense Public Health Database

Francisella tularensis subsp. tularensis Schu 4, complete genome.

Showing 5 kbp from NC_006570, positions 1,432,960 to 1,437,959

Instructions
Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.

Examples: NC_006570:1..20000, Gene:panB, NC_006570.

[Hide banner] [Bookmark this] [Link to Image] [Help] [Reset]

Search

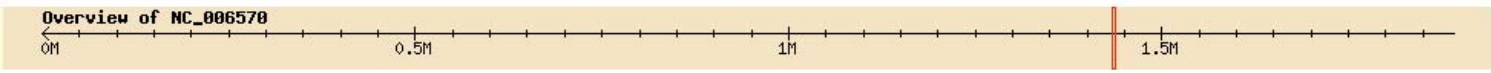
Landmark or Region: NC_006570:1432960..1437959 Search

Data Source: Francisella tularensis subsp. tularensis Schu 4, complete genome.

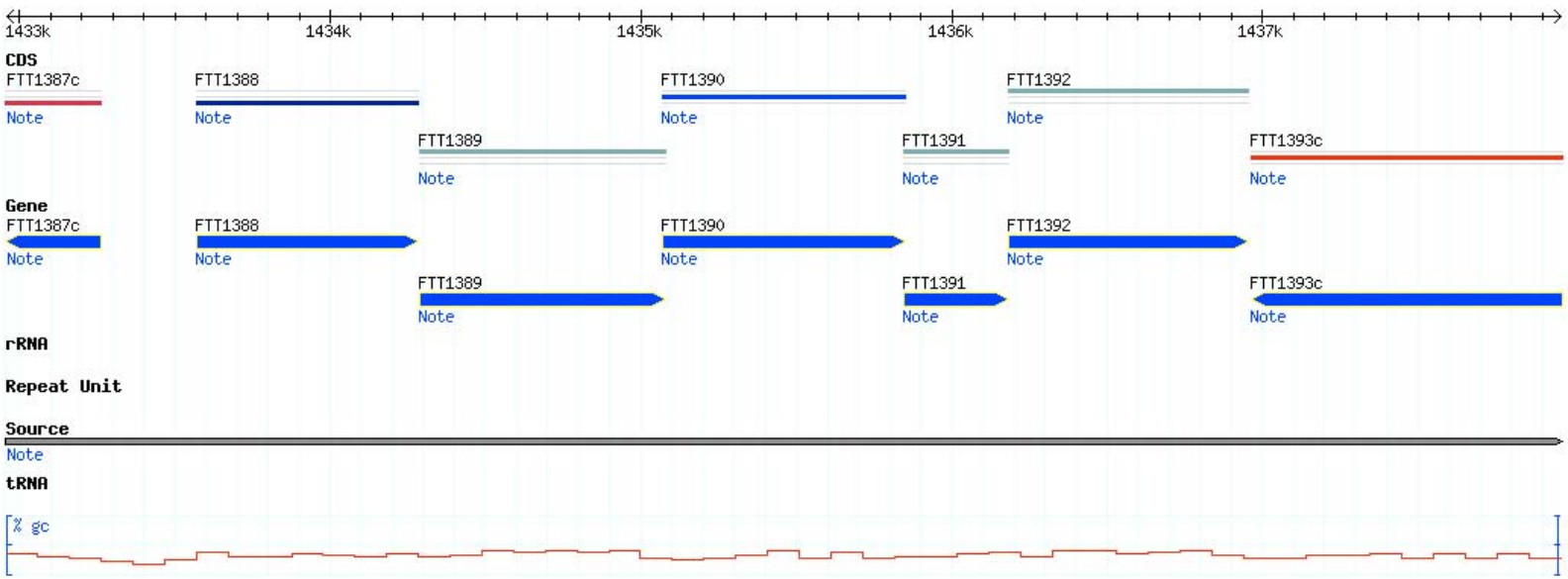
Reports & Analysis: Annotate Restriction Sites Configure... Go

Scroll/Zoom: Show 5 kbp Flip

Overview



Details



Gene Details

Strain Identification *1

Organism Name: Francisella tularensis
Strain Name: Schu 4
NCBI Taxon ID: 177416

Gene Identification *1

Gene Product Name: hypothetical protein
Locus Name: FTT1388
Entrez Gene ID: 3191491
Comment: Similar to AAO81619 (Q833S4) Conserved domain protein from Enterococcus faecalis (281 aa). FASTA: opt: 171 Z-score: 214.0 E(): 0.00046 Smith-Waterman score: 194;26.337 identity in 243 aa overlap ORF ftt1388

Genomic Location *1

Genbank Genome Accession: NC_006570
Coordinates(5'..3'): 1433568 .. 1434284
Strand: Forward
Gene Length: 717
Sequence: View Gene Sequence

Gene Features *1, *2

CDS	Start	End	Protein Length(aa)	Sequence
1	1433568	1434284	238	Protein

Predicted Protein Sequence

BioHealthBase
BioDefense Public Health Database

Protein Sequence

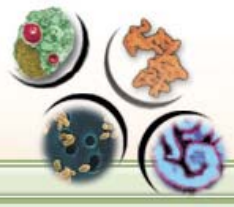
Downloads

GFF submit

Protein Name : Hypothetical protein

Sequence :

```
MQQVPRYVIV GNGNVAAHMC YYFECLKLDF RQWSRNESLD QLDKLLDNAT  050
HVLVLIKDSE IQNFVDRHLT NKSKRLIIIH FSGLLDIKNA YSAHPLQSFP  100
DKNLYSLDEY KSIAFVTCDR SIAFSELLPK LPNANFCIDK SQKAYYHAMC  150
VLANNVSTLI WQKFYTEMQN RFGINQGYLI PFLETTFKNI KHNHHALSGP  200
IARGDNLTLQ KDLDALIGDD FYDVFRAIVN QFSNKEKR               238
```



Blastp Results

BioHealthBase
BioDefense Public Health Database

Sequences producing significant alignments:	Score (Bits)	E Value	
gi|56604929|emb|CAG46021.1| conserved hypothetical protein [F...	489	4e-137	G
gi|62262130|gb|AAX78047.1| unknown protein [synthetic construct]	489	4e-137	
gi|54113333|gb|AAV29300.1| NT02FT2059 [synthetic construct]	452	4e-126	
gi|76796969|ref|ZP_00779315.1| conserved hypothetical protein...	58.2	3e-07	
gi|68195442|gb|EAN09888.1| conserved hypothetical protein [En...	56.6	7e-07	
gi|48856975|ref|ZP_00311132.1| COG5495: Uncharacterized conse...	53.1	8e-06	
gi|73660326|emb|CAI82933.1| conserved hypothetical protein [D...	53.1	8e-06	G
gi|83857084|ref|ZP_00950612.1| hypothetical protein CA2559_09...	53.1	8e-06	
gi|85859369|ref|YP_461571.1| hypothetical cytosolic protein [...	50.4	5e-05	
gi|83589012|ref|YP_429021.1| NADP oxidoreductase, coenzyme F4...	48.5	2e-04	G
gi|67916874|ref|ZP_00510563.1| conserved hypothetical protein...	47.8	3e-04	
gi|71541644|ref|ZP_00663380.1| conserved hypothetical protein...	47.4	4e-04	
gi|85830272|gb|EAQ48732.1| hypothetical protein MED217_09295 [Fl	45.1	0.002	
gi|29542219|gb|AAO91156.1| conserved hypothetical protein [Co...	44.3	0.004	G
gi|57224829|gb|AAW39886.1| pyrroline-5-carboxylate reductase,...	43.9	0.005	G
gi|85819354|gb|EAQ40513.1| hypothetical protein MED134_07149 [Ce	43.5	0.006	
gi|83815088|ref|YP_445037.1| NADP oxidoreductase [Salinibacte...	43.1	0.008	G

Alignment

BioHealthBase
BioDefense Public Health Database

```
 gi|83571926|gb|ABC18478.1|  NADP oxidoreductase, coenzyme F420-dependent [Moorella thermoacetica
ATCC 39073]
Length=307

 Score = 48.5 bits (114),  Expect = 2e-04
 Identities = 45/148 (30%), Positives = 67/148 (45%), Gaps = 9/148 (6%)

Query  90   AYSAHPLQSFPDKNLYSLDEYKSIAFVTCDR-SIAFSE-LLPKLPNANFCIDKSQKAYYH  147
            A + HPLQS  D ++   +   S+  +  DR ++   E L+  L    F I    K  YH
Sbjct  113  ALALHPLQSCADADMAVANLPGSVFSLEGDREALPLGERLVNDLEGEYFIISPEAKPLYH  172

Query  148  AMCVLANNVSTLIWQKFYTEMQNRF---GINQGYLIPFLETTFKNIKHNH--HALSGPIA  202
            A   +A+N    I    Y  MQ       +    L P +E T+ NIK      AL+GPI
Sbjct  173  AAACVASNYLVSIVDLSYRLMQAAGMAPDMVARALAPLIEGTWGNIKEKGVPRALTGPIT  232

Query  203  RGDNLTLQKDLDALI--GDDFYDVFRAI  228
            RGD  T+   L A+     +  +++RA+
Sbjct  233  RGDVATIASHLQAMAARAPELEEIYRAV  260


>gi|67916874|ref|ZP_00510563.1|  conserved hypothetical protein [Clostridium thermocellum ATCC
27405]
 gi|67849181|gb|EAM44803.1|  conserved hypothetical protein [Clostridium thermocellum ATCC
27405]
Length=289

 Score = 47.8 bits (112),  Expect = 3e-04
 Identities = 51/204 (25%), Positives = 94/204 (46%), Gaps = 33/204 (16%)

Query  40   DQLDKLLDNATHVLVLIKDS---EIQNFVDRHLTNKSKRL-IIIHFSGLLDIK-------  88
            + L+  + N+  + + + D+   E+   + R + N + R    IH SG L  K
Sbjct  50   NDLEDAVKNSDVIFISVSDNNIIEVAEEIVRKVDNDALRSKTFIHMSGALTAKALKPLEN  109

Query  89   -NAY--SAHPLQSFPDKN-----LYSLDEYKSIAFVTCDRSIAFSELLPKLPNANFCIDK  140
              AY  S HP+QS  DK+     LY++       F  C+ ++  +  + K          K
Sbjct  110  LGAYTGSLHPIQSVADKDSGWKKLYNI----YYGFEGCEEALKHALTVVKSFEGKLIKIK  165

Query  141  SQ-KAYYHAM-CVLANNVSTLIWQKFYTEMQNRFGINQGY----LIPFLETTFKNIKH--  192
             Q K  YHA  C+++N   TL +  +  ++    G ++         +P ++ T  NI+
Sbjct  166  EQDKTLYHAAACIISNYTVTLSYVAY--KILESIGFDRETADKAFLPLIKNTVHNIERLG  223

Query  193  NHHALSGPIARGDNLTLQKDLDAL  216
```

Semi-Automated Hole Filling

BioHealthBase
BioDefense Public Health Database

- Identify pathway holes
- Identify hypothetical proteins linked to characterized pathway components through operon predictions
- Blastp against NR database
- Multiple sequence alignment
- Curate conserved regions for catalytic residues
- Annotate record with proposed function
- Add hypothesis track to Gbrowse

Influenza Home Page

Influenza Virus

BioHealthBase

HOME | FEEDBACK | OUR MISSION | HELP

Database Search
- Gene
- Epitope
- Public Database Identifier
- Sequence Polymorphism

Reactome Pathway
- Influenza Infection

Release Notes
FAQ
Feedback
Related Links
Data Loads

What's New

- Nov 3, 2005 : Began collaboration with Reactome team on interferon pathway analysis
- Jan 31, 2006 : First production BioHealthBase release
- Feb 1, 2006 : Reactome release with Influenza lifecycle and interferon interference additions
- Mar 28, 2006 - Apr 2, 2006 : Keystone conference
  - Advances in Influenza Research: From Birds to Bench to Bedside (X8)
  - Sheraton Steamboat Resort
  - Steamboat Springs, Colorado

About *Influenza Virus*

Influenza virus, a member of the Orthomyxoviridae family, consists of a segmented, negative-stranded RNA genome encoding ten proteins. The Influenza virus species is subdivided into three subtypes A, B and C.

The Influenza viruses' life cycle may be broken down into six stages: binding to the host, entry into the host, fusion and uncoating, nuclear import of ribonucleoproteins, nuclear export of ribonucleoproteins and virus assembly and release.

The Influenza viruses bind via their surface HA (hemagglutinin) to a host cell's sialic acid in alpha 2,3 or alpha 2,6 linkage. Sialic acids in human cells contain 2,6 linkages while 2,3 linkages are found in avian cells. The specificity of influenza HA for sialic acid in alpha 2,6 or alpha 2,3 linkages normally prevents the transfer of influenza viruses between avian species and humans. This restriction can be overcome, however as observed in viruses in culture which adapt to their host though mutation in the receptor-binding site of the viral HA gene.

Once bound virus particles can be internalized by four mechanisms.The majority of internalization events appear to be mediated by clathrin-coated pits, but internalization via caveolae, macropinocytosis, and by non-clathrin, non-caveolae pathways has also been described for influenza viruses.

Genome Statistics

General Information for *Influenza Virus*

Genome Type	segmented (-) single-stranded RNA
Genome Size	~ 12-14 kb
Replication	Nuclear
Taxon ID	11308
Kingdom	Virus

***Influenza* Genomes in BioHealthBase**

Species	# Strain	# Host	# Subtype
Influenza A Virus	5102	16	105
Influenza B Virus	1040	3	0
Influenza C Virus	152	2	0
Total	6294	21	105

Genome Statistics for Reference Strain A/PR/8/34

Segment	Length (nt)	Encoded Protein	Protein Length (aa)
1	2341	PB2	759
2	2341	PB1	757
3	2233	PA	716
4	1778	HA	566
5	1565	NP	498
6	1413	NA	454
7	1027	M1; M2	252; 97
8	890	NS1; NS2	230; 121

Gene Query

BioHealthBase
BioDefense Public Health Database

Gene Search

Text search of genes in BioHealthBase. To refine your search, specify as many fields as possible. ?

Search By Organism: Influenza A Virus

Subtype : H5N1 * eg: H1N1 (Leave blank to view genes from all subtypes)

Segment :
All
1
2
3
4
5
6
7
8

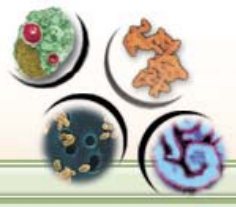
From Year : 2002 eg. 1999

To Year : eg. 2005

Host : All

Geographic Location: All

Go Clear



Gene Summary Page

BioHealthBase
BioDefense Public Health Database

Influenza Gene Search Results

Your search returned " **289** " records

New Search

1 | 2 | 3 |

GFF submit

ALL	Gene Symbol	Gene Product Name	Segment	Genbank Genome Accession	UniProtKB Accession	Organism	Geographic Location	Year	Host
☐	HA	hemagglutinin	4	AY590569	Q6PNF7	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*	Thailand	2004	Avian
☐	HA	hemagglutinin	4	AY646175	Q6DTX2	*Influenza A Virus A/leopard/Suphanburi/Thailand/Leo-1/04*	Thailand	2004	Leopard
☐	HA	hemagglutinin	4	AY590571	Q6PNF5	*Influenza A Virus A/leopard/Thailand/CU-MD/2004*	Thailand	2004	Leopard
☐	HA	hemagglutinin	4	AY535021	Q6QLM5	*Influenza A Virus A/leopard/Thailand/TLV3/2004*	Thailand	2004	Leopard
☐	HA	hemagglutinin	4	AY535020	Q6QLM6	*Influenza A Virus A/leopard/Thailand/TM3/2004*	Thailand	2004	Leopard
☐	HA	hemagglutinin	4	AY553802	Q6Q7T5	*Influenza A Virus A/little grebe/Thailand/Phichit-01/2004*	Thailand	2004	Avian
☐	HA	hemagglutinin	4	DQ017280	Q4U3A9	*Influenza A Virus A/littlecuckoo-dove/Tak-2-01/2004*	Thailand	2004	Avian
☐	HA	hemagglutinin	4	AY553803	Q6Q7T4	*Influenza A Virus A/muscovy duck/Thailand/Tak-01/2004*	Thailand	2004	Avian
☐	HA	hemagglutinin	4	AY576930	Q6PUS9	*Influenza A Virus A/muscovy duck/Vietnam/MdGL/2004*	Viet Nam	2004	Avian
☐	HA	hemagglutinin	4	AY590577	Q6PNE9	*Influenza A Virus A/openbill/Thailand/CU-2/2004*	Thailand	2004	Avian
☐	HA	hemagglutinin	4	DQ083574	Q45ZQ4	*Influenza A Virus A/ostrich/Samut Prakan/Thailand/CU-31/04*	Thailand	2004	Avian
☐	HA	hemagglutinin	4	AY585368	Q6E420	*Influenza A Virus A/duck/Shanghai/35/2002*	China	2002	Avian
☐	HA	hemagglutinin	4	AY585369	Q6E419	*Influenza A Virus A/duck/Shanghai/37/2002*	China	2002	Avian
☐	HA	hemagglutinin	4	AY779048	Q5U8M5	*Influenza A Virus A/duck/Thailand/CU-2/2004*	Thailand	2004	Avian
☐	HA	hemagglutinin	4	AY553797	Q6Q7U0	*Influenza A Virus A/duck/Thailand/Kamphaengphet-01/2004*	Thailand	2004	Avian

Gene Detail Page

BioHealthBase
BioDefense Public Health Database

Gene Details

Strain Identification *1

Organism Name:	Influenza A Virus
Strain Name:	A/kalij pheasant/Thailand/CU-4/2004
Subtype:	H5N1
Host:	Avian
Date Isolated:	2004
Country:	Thailand
NCBI Taxon ID:	270486

Gene Identification *1

Gene Symbol:	HA
Gene Product Name:	hemagglutinin
Comment:	

Genomic Location *1

Segment:	4
Genbank Genome Accession:	AY590569
Coordinates(5'..3'):	1 .. 1626
Gene Length:	1626
Sequence:	View Gene Sequence

Gene Features *1, *2

CDS	Start	End	Protein Length(aa)	Sequence
1	1	1626	542	Protein

Protein Identification *2

Top

Protein Name:	Hemagglutinin
UniProtKB Accession:	Q6PNF7
Genbank Protein Accession:	AAS89269.1
Genbank Protein GI:	46361438
Comment:	
Keywords:	Envelope protein;Hemagglutinin

HMM/Pfam Domains *4

Top

Accession	Name	Description	Start	End
PF00509.7	Hemagglutinin	Hemagglutinin	19	541

Motifs *4

Top

Motif	Start	End	Program
coiled_coil	402	436	ncoils
sig_p	1	16	signalp

Epitope Prediction

Top

Epitope Details

Gene Ontology *2

Top

Type	Name	Go ID
COMPONENT	viral envelope	GO:0019031
PROCESS	heterophilic cell adhesion	GO:0007157
PROCESS	viral infectious cycle	GO:0019058

Database Cross References *2

Top

Database Name	Accession	Description
SMR	Q6PNF7	N/A
INTERPRO	IPR001364	Hemagglutn

Links & Data Sources

BioHealthBase
BioDefense Public Health Database

Database Cross References *2

Top

Database Name	Accession	Description
SMR	Q6PNF7	N/A
INTERPRO	IPR001364	Hemagglutn
INTERPRO	IPR000149	Hemagglutn_1
PFAM	PF00509	Hemagglutinin
PRINTS	PR00330	HEMAGGLUTN1
PRINTS	PR00329	HEMAGGLUTN12
PRODOM	PD000225	Hemagglutn

References *1, *2

Top

PubMed ID	Journal Name	Title	Author

* Data Sources

Top

ID	Source
1	NCBI
2	UniProtKB
4	Pfam

Epitope Prediction

BioHealthBase
BioDefense Public Health Database

Epitope Search

This page allows you to search for epitope predictions in BioHealthBase. We have used NETCTL 1.0 to predict epitopes in Influenza protein sequences.To learn more about epitopes and MHC Superfamily, reference our Help document.

To refine your search, specify as many text fields and drop down menus as possible. ?

Search By Strain (OR) **Search By Identifier**

Strain Name: A/kalij pheasant/Thailand,

Select Segment(s) : All 1 2 3 4 5 6 7 8

MHC Superfamily : All A2 A3 A24 B7 B44

Public Database Identifier: Genbank Protein Accession

Identifier Keyword:

From AA :

To AA :

Score Cutoff :

* Specify start and end field to view predictions from a given protein range.

* Specify this field to view predictions above a cutoff score.

Go Clear

Predicted Epitopes

BioHealthBase
BioDefense Public Health Database

Influenza Epitope Search Results

Your search returned " **14** " records

New Search

Gene Symbol	Gene Product Name	Genbank Protein Accession	Genbank Protein GI	UniProtKB Accession	Peptide Sequence	Score	AA Start	MHC Superfamily	Organism
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	FFRNVVWLI	0.7793	159	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	TIMEKNVTV	0.97	34	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	GLFGAIAGF	1.2306	347	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	AIDGVTNKV	1.2575	390	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	KMNTQFEAV	0.7784	404	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	KMEDGFLDV	1.3914	429	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	FLDVWTYNA	1.3018	434	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	VLMENERTL	0.9845	446	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	QLRDNAKEL	0.7803	471	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	VLLLAIVSL	0.7891	5	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	LLLAIVSLV	1.0618	6	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	PMCDEFITV	1.1019	81	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	TVPEWSYIV	1.1512	88	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*
HA	hemagglutinin	AAS89269.1	46361438	Q6PNF7	YIVEKANPV	1.138	94	A2	*Influenza A Virus A/kalij pheasant/Thailand/CU-4/2004*

Subtype Polymorphism

BioHealthBase
BioDefense Public Health Database

Sequence Polymorphism Search

To study the sequence polymorphism across different influenza strains, we downloaded all influenza sequences from Genbank. We then generated multiple alignments on sequences with the same subtype using the **MUSCLE** multiple alignment tool.

A consensus sequence for each subtype was created by following the majority rule. For each position in the multiple alignment, a score was generated by using a formula modified from the one as described in **Crooks et al.**. The score ranges from 0 (no polymorphism) to 200 (highest polymorphism). For more details on the score and general polymorphism approach, reference our **Help** document.

To view a polymorphism table, select the subtype and the segment, then click "Go" button. ?

Select Subtype: H5N1

Select Segment: 4

Go Clear

Allele Count & Score

BioHealthBase
BioDefense Public Health Database

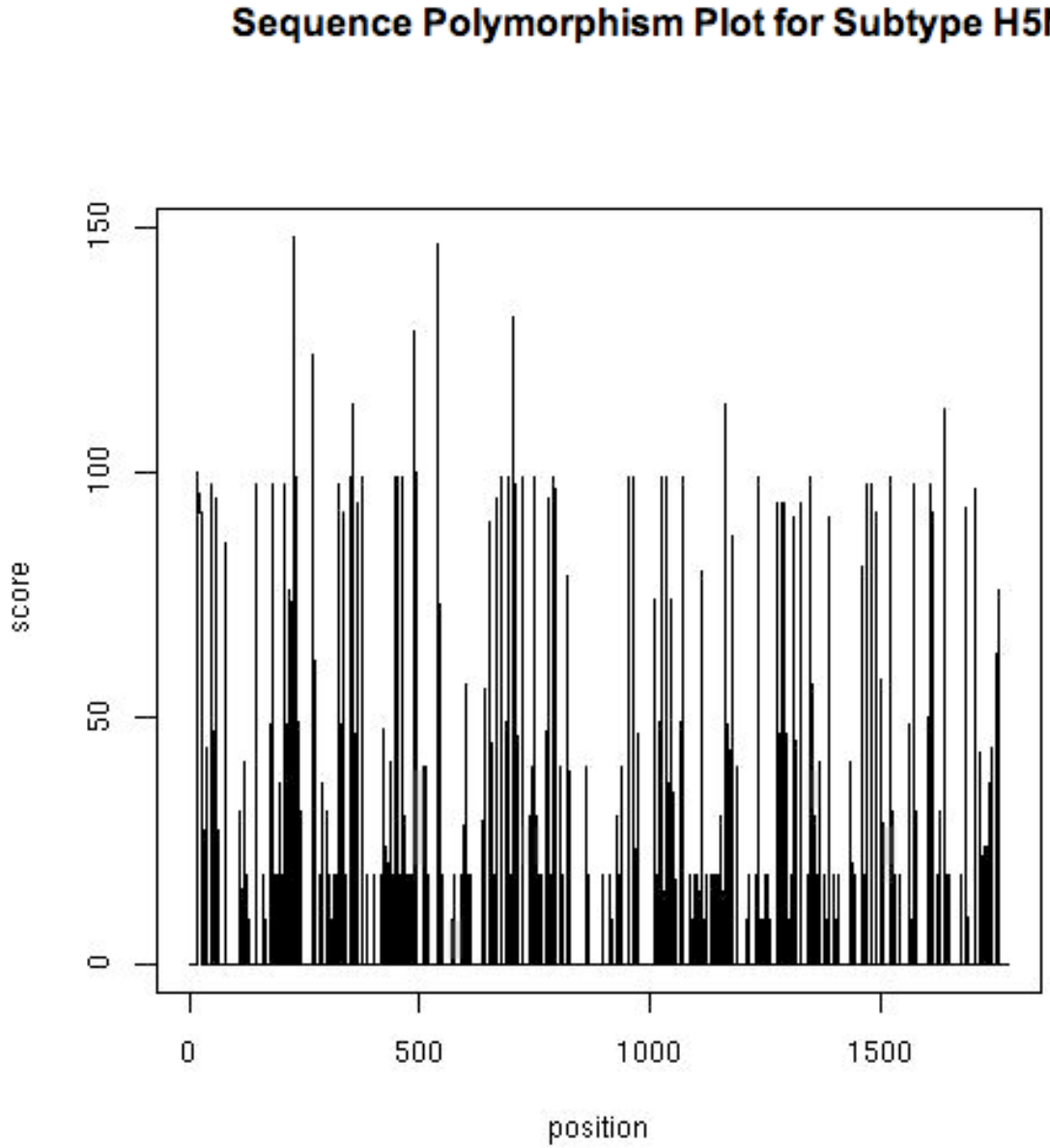
Sequence Polymorphism Analysis for Subtype H5N1, Segment 4

Figure View
Raw Alignment

New Search

Position	Coding	Score	Consensus	A	T	G	C	Deletion	# Sequences
1	no	0	A	2	0	0	0	0	2
2	no	0	G	0	0	2	0	0	2
3	no	0	C	0	0	0	2	0	2
4	no	0	A	2	0	0	0	0	2
5	no	0	A	2	0	0	0	0	2
6	no	0	A	2	0	0	0	0	2
7	no	0	A	2	0	0	0	0	2
8	no	0	G	0	0	2	0	0	2
9	no	0	C	0	0	0	2	0	2
10	no	0	A	2	0	0	0	0	2
11	no	0	G	0	0	2	0	0	2
12	no	0	G	0	0	2	0	0	2
13	no	0	G	0	0	2	0	0	2
14	no	0	G	0	0	2	0	0	2
15	no	0	T	0	2	0	0	0	2
16	no	0	T	0	2	0	0	0	2
17	no	0	C	0	0	0	2	0	2
18	no	100	N	3	3	0	0	0	6
19	no	100	N	3	0	0	3	0	6
20	no	100	N	3	3	0	0	0	6
21	no	92	C	0	3	0	6	0	9
22	no	92	T	0	6	0	3	0	9
23	no	92	G	0	3	6	0	0	9
24	no	92	T	0	6	3	0	0	9
25	no	92	C	0	3	0	6	0	9
26	no	92	A	6	0	0	3	0	9
27	no	0	A	9	0	0	0	0	9
28	no	0	A	9	0	0	0	0	9
29	yes	0	A	22	0	0	0	0	22
30	yes	0	T	0	22	0	0	0	22

Polymorphism Structure



Host-Pathogen Interactions

- What it is -
  - Existing GO term
    - id: GO:0000005
    - name: interaction with host organism
    - def: "Any interaction between an organism, usually a parasite or symbiont, and another organism from which it may obtain nourishment, protection, and/or a means of dispersal." [PAMGO:cc ""]
    - is_a: GO:0000001
  - New GO term
    - name: pathogenic interaction with host organism
    - def: "Any interaction between an organism, and another host organism from which it may obtain nourishment, protection, and/or a means of dispersal *to the detriment of the host*."
- Why it is important to emphasize
  - Understanding of the epidemiology of transmission
  - Impacts host range
  - Impacts virulence
  - Drug and vaccine targets
- Different meanings at different levels of granularity
  - Ecological
  - Physiological
  - Cellular
  - Molecular

Host-Pathogen Interactions - Viruses

- Cell intrinsic response modulators
  - Subvert metabolic machinery (e.g. influenza mRNA cap stealing)
  - Counteract cell death (e.g. herpesvirus bcl2 family members)
  - Subvert IRF activation (e.g. Influenza virus NS1)
- Innate/inflammatory response modulators
  - Counteract interferon action (e.g. Henipaviruses V, W, P proteins)
  - Modulate/evade NK cell recognition
  - Cytokine/chemokine control system (herpesvirus chemokine-like factor)
- Adaptive immune response modulators
  - Modulate/evade antibody activity (e.g. serologic evolution of influenza)
  - Modulate/evade TH response (isolate evolution of influenza)
  - Modulate/evade CTL response (isolate evolution of influenza)

Rationale for Flu & Reactome

BioHealthBase
BioDefense Public Health Database

- Useful features of Reactome
  - Framework in place
  - Focus on human biological processes that are frequently targeted by pathogens
  - Emphasize assertions specifically supported in literature
  - SkyPainter for conveying attribute information
- Approach for Flu Host-Pathogen Pathways
  - Step 1 - High level framework of viral life cycle
  - Step 2 - Detailed host pathways
  - Step 3 - Detailed viral pathways
  - Step 4 - Host-viral pathway intersection

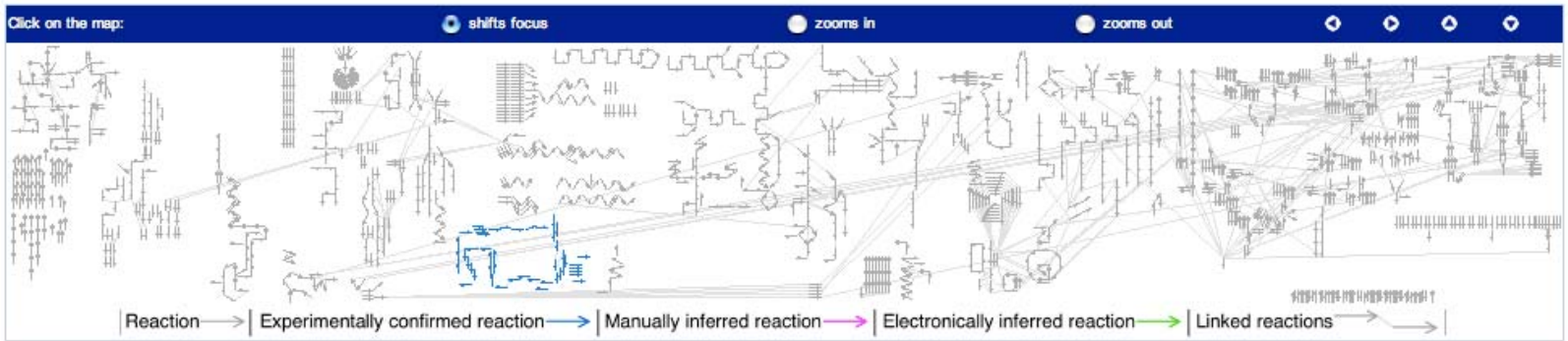
Influenza Life Cycle in Reactome

BioHealthBase
BioDefense Public Health Database

Home | About | TOC | User Guide | Data Model | Schema | Extended Search | PathFinder | SkyPainter | Download | Linking | Citing | Editorial Calendar | Help

Find everything with ALL of the words in Homo sapiens Go!

Reactionmap



Event hierarchy

open to selected event | open all | close all | show/hide hierarchy types

Influenza Infection [Homo sapiens]

Details

Influenza Infection

Luo, F, Squires, B, Scheuermann, RH, 2006-01-05

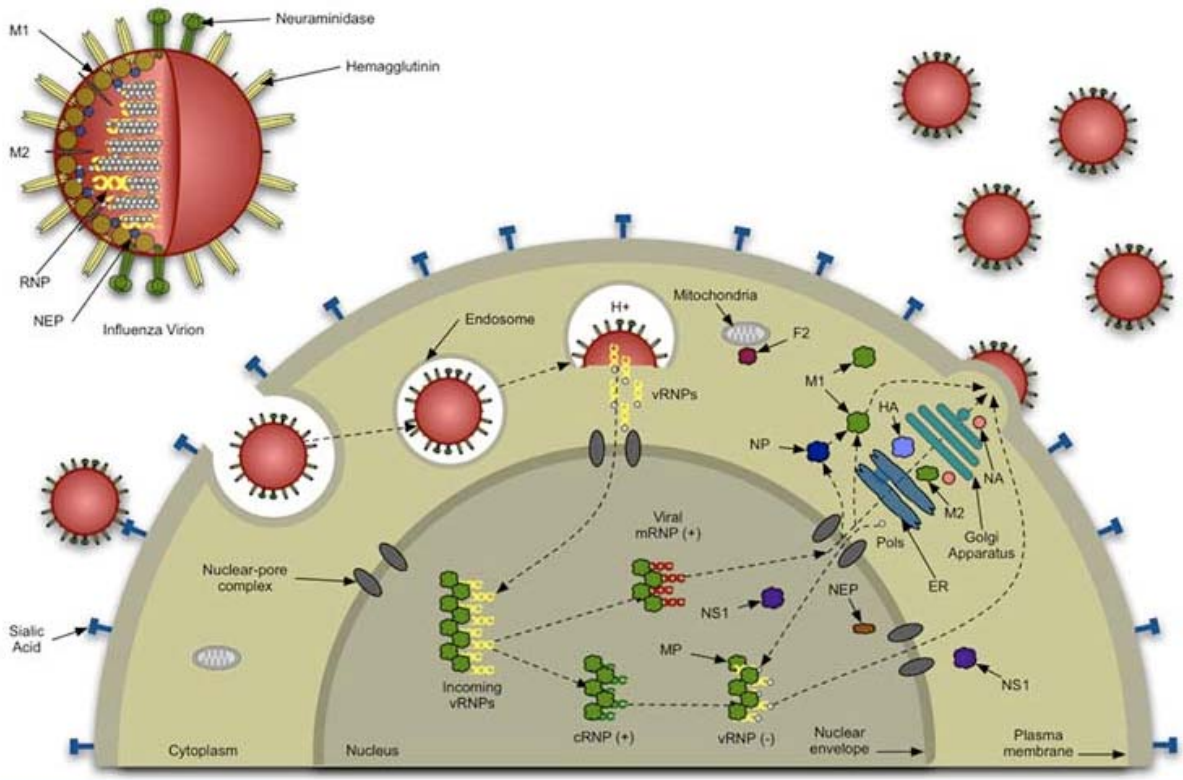
For centuries influenza epidemics have plagued man, and influenza was probably the disease described by Hippocrates in 412 BC. Today it remains a major cause of morbidity and mortality worldwide with large segments of the human population affected every year. Many animal species can be infected by influenza viruses, often with catastrophic consequences. A continuing threat is the possibility of a pandemic similar to that experienced in 1918, estimated to have been responsible for 50 million deaths worldwide.

Influenza viruses belong to the family of Orthomyxoviridae; viruses with segmented RNA genomes that are negative sense and single-stranded (Baltimore 1971).

Influenza virus strains are named according to their type (A, B, or C), the species from which the virus was isolated (omitted if human), location of isolate, the number of the isolate, the year of isolation, and in the case of influenza A viruses, the hemagglutinin (H) and neuraminidase (N) subtype. For example, the virus of H5N1 subtype isolated from chickens in Hong Kong in 1997 is: influenza A/chicken/Hong Kong/220/97(H5N1) virus. Currently 16 different hemagglutinin (H1 to H16) subtypes and 9 different neuraminidase (N1 to N9) subtypes are known for influenza A viruses. Most human disease is due to Influenza viruses of the A type, so the events of Influenza infection have been annotated in Reactome with reference to this type. [Krug & Lamb 2001]

Life Cycle Schematic

BioHealthBase
BioDefense Public Health Database



Organism	Homo sapiens Influenza A virus

References

Krug, RM, Lamb, RA *Orthomyxoviridae: The Viruses and Their Replication* **2001** Fields Virology. 4th edition, editors: Knipe DM, Howley PM,. Philadelphia: Lippincott Williams & Wilkins. ISBN: 0-7817-1832-5 [Orthomyxoviridae: The Viruses and Their Replication]

High Level Pathways

BioHealthBase
BioDefense Public Health Database

- Influenza Infection [Homo sapiens]
  - Influenza Life Cycle
    - Binding of the influenza virion to the host cell
    - Entry of Influenza Virion into Host Cell via Endocytosis
      - Clathrin-Mediated Endocytosis of the Influenza Virion
      - Caveolae Endocytosis
      - Non-Clathrin, Non-Caveolae Endocytosis Pathways
      - Macropinocytosis
    - Fusion and Uncoating of the Influenza Virion
      - Fusion of the Influenza Virion to the Host Cell Endosome
      - Uncoating of the Influenza Virion
    - Transport of Ribonucleoproteins into the Host Nucleus
      - Recognition of the Nuclear Localization Signal (NLS) by a Karyopherin Alpha Family Protein
      - Recruitment of Karyopherin Beta to form a Trimeric Complex
      - Docking of trimeric complex at the nuclear pore
      - Transport of the RNP into the host cell nucleus
      - Release of the RNP into the host cell nucleus
    - Influenza Virion RNA Transcription and Replication
      - Viral Messenger RNA Synthesis
      - cRNA Synthesis
      - vRNA Synthesis
    - Assembly and Nuclear Export of Ribonucleoproteins
      - Assembly of Viral RNP Complexes in the Host Cell Nucleus
      - NEP/NS2 Interacts with the Cellular Export Machinery
    - Virus Assembly and Release
      - Assembly of Viral Components at the Budding Site
      - Packaging of Eight RNA Segments
      - Budding
      - Release
  - Host Interactions with Influenza Factors
    - NS1 Mediated Effects on Host Pathways
    - Influenza Virus Induced Apoptosis

- Assembly and Nuclear Export of Ribonucleoproteins
  - **Assembly of Viral RNP Complexes in the Host Cell Nucleus [Homo sapiens]**
  - NEP/NS2 Interacts with the Cellular Export Machinery
- Virus Assembly and Release
  - Assembly of Viral Components at the Budding Site
  - Packaging of Eight RNA Segments
  - Budding
  - Release
- Host Interactions with Influenza Factors
  - NS1 Mediated Effects on Host Pathways
  - Influenza Virus Induced Apoptosis

Details

Assembly of Viral RNP Complexes in the Host Cell Nucleus	
Viral RNP is assembled in the host cell nucleus through the interaction of full-length negative strand viral RNA (vRNA) and the viral nucleocapsid (NP) and matrix (M1) proteins. Studies of interactions of the purified components in vitro and of tissue culture model systems expressing various combinations of the components have established roles for both NP and M1 proteins in the assembly of a complex that has the physical properties of RNP purified from virions and that can be exported from the host cell nucleus (Huang et al. 2001; Whittaker et al. 1996). [Huang *et al* 2001, Whittaker *et al* 1996]	
Following event(s)	NEP/NS2 Interacts with the Cellular Export Machinery [Homo sapiens]
Organism	Homo sapiens Influenza A virus
Cellular compartment	nucleoplasm [GO:0005654]
References	
Huang, X, Liu, T, Muller, J, Levandowski, RA, Ye, Z *Effect of influenza virus matrix protein and viral RNA on ribonucleoprotein formation and nuclear export* **2001** Virology [11531417] Whittaker, G, Bui, M, Helenius, A *Nuclear trafficking of influenza virus ribonuleoproteins in heterokaryons* **1996** J Virol [8627748]	
Represents GO biological process	viral genome maturation [GO:0019070]

[Change default viewing format] [SBML] [BioPAX] [PDF] [Cytoscape] [Protégé] List...

Host Response Pathways

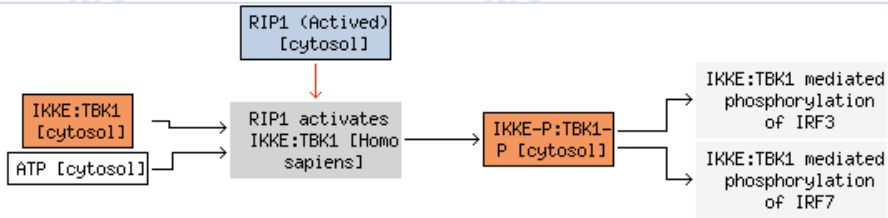
BioHealthBase
BioDefense Public Health Database

- Signaling Pathways
  - Glucagon signalling in metabolic regulation
  - Opioid Signalling
  - NGF signaling
  - NMDA-mediated neuronal signaling
  - Notch Signaling Pathway
  - Rho GTPases
  - Immune System Signalling
    - Adaptive Immunity Signaling
    - Innate Immunity Signaling
      - Complement cascade
      - **Induction of the Antiviral Cytokine IFN-alpha/beta Cascades [Homo sapiens]**
        - Nuclotide-binding Oligomerization Domain (NOD) Signaling Pathway
        - RIG-I Mediated Induction of the Antiviral Cytokine IFN-alpha/beta Cascades
        - TRAF6 Mediated Induction of the antiviral cytokine IFN-alpha/beta cascade
        - Toll Receptor Cascades
      - Toll Receptor Cascades
        - Toll Like Receptor 10 (TLR10) Cascade
        - Toll Like Receptor 3 (TLR3) Cascade
          - Viral dsRNA bound to TLR3
          - Association of TRAM to activated TRL3
          - Assoaciation of TRIF to TRAM compled to acitvated TRL3
          - TRIF:TRAM:TLR3 complex recruit RIP1
          - TRIF:TRAM:TLR3 complex activate RIP1
          - RIP1 Mediated Induction of the Antiviral Cytokine IFN-alpha/beta Cascade
            - RIP1 active IRF3/7 and induce Interferon A/B Cascade
              - RIP1 activates IKKE:TBK1
              - IKKE:TBK1 mediated phosphorylation of IRF3
              - IKKE:TBK1 mediated phosphorylation of IRF7
              - Transcriptional action of IRF3 induce IFNA
              - Transcriptional action of IRF3 induce IFNB
              - Transcriptional action of IRF7 induce IFNB
              - Transcriptional action of IRF7 induce IFNA
            - RIP1 activates NFkB and induces Interferon A/B Cascade
        - Toll Like Receptor 5 (TLR5) Cascade
        - Toll Like Receptor 7 (TLR7) Cascade
        - Toll Like Receptor 8 (TLR8) Cascade

Host Response Detail

- RIP1 Mediated Induction of the Antiviral Cytokine IFN-alpha/beta Cascade
  - RIP1 active IRF3/7 and induce Interferon A/B Cascade
    - **RIP1 activates IKKE:TBK1 [Homo sapiens]**
    - IKKE:TBK1 mediated phosphorylation of IRF3
    - IKKE:TBK1 mediated phosphorylation of IRF7
    - Transcriptional action of IRF3 induce IFNA
    - Transcriptional action of IRF3 induce IFNB
    - Transcriptional action of IRF7 induce IFNB
    - Transcriptional action of IRF7 induce IFNA
  - RIP1 activates NFkB and induces Interferon A/B Cascade
- Cessation Of Protein Translation Due To The Phosphorlation Of The eIF2 Complex By PKR
- Insulin receptor mediated signalling

Diagram



Details

RIP1 activates IKKE:TBK1	
Luo, F, 2005-11-10	
Input	IKKE:TBK1 [cytosol] ATP [cytosol] C
Output	IKKE-P:TBK1-P [cytosol]
Catalyst	RIP1 (Actived) [cytosol] U
GO molecular function	protein kinase activity [GO:0004672]
Following event(s)	IKKE:TBK1 mediated phosphorylation of IRF3 [Homo sapiens] IKKE:TBK1 mediated phosphorylation of IRF7 [Homo sapiens]
Organism	Homo sapiens
Participating molecules	

- ATP [cytosol] C
- IKKE [cytosol]
- IKKE-P [cytosol]
- RIP1 (Actived) [cytosol] U

Future Reactome enhancements

BioHealthBase
BioDefense Public Health Database

- Engage flu community to assist in pathway building
- Drill down life cycle stages
- Handle functional sequence polymorphisms
- Capture context-dependence in space (cell-type specificity; host range), time (kinetics) and state of host (variations in immune system function)

Summary and Future Directions

BioHealthBase
BioDefense Public Health Database

- Semi-automated Data Loading
  - Added, updated, and deleted sequences and annotation
  - Added, updated, and deleted protein information in UniProt, PFam, etc.
  - “Catchup” loads plus regularly-scheduled loads
- Additional Tools
  - User-interactive sequence alignment and multi-alignment
  - Domain, GO, motif searches
  - Enhanced polymorphisms displays
- Science Collaborations
  - UTSW Francisella researchers
  - UCI Francisella microarrays
  - IEDB validated epitopes
  - Reactome pathway analysis
- Additional pathogens
- Outreach

Acknowledgments

- U.T. Southwestern
  - Burke Squires
  - Feng Luo
  - Shubhada Godbole
  - Jennifer Cai
  - Jyoti Shah
  - Jamie Lee
  - Cathy Spranger
  - Victoria Hunt
  - Kevin McIver
- SRI
  - Peter Karp
  - John Pick
- Co-investigators
  - Barbara Mann (UVA)
  - Adolfo Garcia-Sastre (MSSM)
  - Stephen Johnston (UTSW)
  - Hillary Morrison (MBL)
  - Louis Weiss (Albert Einstein)
  - Ellen Vitetta (UTSW)

•Northrop Grumman
  –Carey Gire
  –Kevin Biersack
  –Jason Lucas
  –Sharmila Vattikuti
  –Sanjeev Kumar
  –Paul Shrabstein
  –Surabhi Sharma
  –Tammie Ajayi
  –Aihui Wang
  –Zuoming Deng
  –Jianjun Wang
  –DeWayne Ejikeme
  –Doug Marcey

•Vecna
  –James Wolowicz
  –Chris Larsen
  –Al Ramsey

Amar
  X. Wei

Reactome
  Marc Gillespie
  Peter D-Eustachio